

```

1 TCGCGGCGGC CGTGATGGCT GGTGACGGCG GGGCCGGGCA GGGGACCGGG
51 GCCGCGGCCC GGGAGCGGGC CAGCTGCCGG GAGCCCTGAA TCACCGCCTG
101 GCCCGACTCC ACCATGAACG TCGCGCTGCA GGAGCTGGGA GCTGGCAGCA
151 ACATGGTGGG GTACAAACGG GCCACGCTTC GGGATGAAGA CGCACCCGAG
201 ACCCCCGTAG AGGGCGGGGC CTCCCCGGAC GCCATGGAGG TGGGCAAGGG
251 GGCTTCCCCC TTCTCACCAG GCCCCAGCCC TGGCATGACG CCTGGCACAC
301 CCAGGAGCTC TGGGCTGTTC TGGAGGGTCA CCTGCCCCCA CCTCCGCTCC
351 ATCTCTGGCC TCTGCTCTAG GACTATGGTG GGATTCCAGA AGGGGACAAG
401 ACAGCTGTTA GGCTCACGCA CGCAGCTGGA GCTGGTCTTA GCAGGTGCCT
451 CTCTACTGCT GGCTGCACTG CTTCTGGGCT GCCTTGTGGC CCTAGGGGTC
501 CAGTACCACA GAGACCCATC CCACAGCACC TGCCTTACAG AGGCCTGCAT
551 TCGAGTGGCT GGAAAAATCC TGGAGTCCCT GGACCGAGGG GTGAGCCCCT
601 GTGAGGACTT TTACCAGTTC TCCTGTGGGG GCTGGATTCTG GAGGAACCCC
651 CTGCCCAGTG GCGGTTCTCG CTGGAACACC TTCAACAGCC TCTGGGACCA
701 AAACCAGGCC ATACTGAAGC ACCTGCTTGA AAACACCACC TTCAACTCCA
751 GCAGTGAAGC TGAGCAGAAG ACACAGCGCT TCTACCTATC TTGCCTACAG
801 GTGGAGCGCA TTGAGGAGCT GGGAGCCCAG CCACTGAGAG ACCTCATTGA
851 GAAGATTGGT GGTTGGAACA TTACGGGGCC CTGGGACCAG GACAACCTTA
901 TGGAGGTGTT GAAGGCAGTA GCAGGGACCT ACAGGGCCAC CCCATTCTTC
951 ACCGTCTACA TCAGTGCCGA CTCTAAGAGT TCCAACAGCA ATGTTATCCA
1001 GGTGGACCAG TCTGGGCTCT TTCTGCCCTC TCGGGATTAC TACTTAAACA
1051 GAACTGCCAA TGAGAAAGTG CTCACTGCCT ATCTGGATTA CATGGAGGAA
1101 CTGGGGATGC TGCTGGGTGG GCGGCCACC TCCACGAGGG AGCAGATGCA
1151 GCAGGTGCTG GAGTTGGAGA TACAGCTGGC CAACATCACA GTGCCCCAGG
1201 ACCAGCGGCG CGACGAGGAG AAGATCTACC ACAAGATGAG CATTTCGAG
1251 CTGCAGGCTC TGGCGCCCTC CATGGACTGG CTTGAGTTCC TGTCTTCTT
1301 GCTGTACCCA TTGGAGTTGA GTGACTCTGA GCCTGTGGTG GTGTATGGGA
1351 TGGATTATTT GCAGCAGGTG TCAGAGCTCA TCAACCGCAC GGAACCAAGC
1401 ATCCTGAACA ATTACCTGAT CTGGAACCTG GTGCAAAAGA CAACCTCAAG
1451 CCTGGACCGA CGCTTTGAGT CTGCACAAGA GAAGCTGCTG GAGACCTCT
1501 ATGGCACTAA GAAGTCTGT GTGCCGAGGT GGCAGACCTG CATCTCCAAC
1551 ACGGATGACG CCCTTGGCTT TGCTTTGGGG TCCCTCTTCG TGAAGGCCAC
1601 GTTTGACCGG CAAAGCAAAG AAATTGCAGA GGGGATGATC AGCGAAATCC
1651 GGACCGCATT TGAGGAGGCC CTGGGACAGC TGGTTTGGAT GGATGAGAAG
1701 ACCCGCCAGG CAGCCAAGGA GAAAGCAGAT GCCATCTATG ATATGATTGG
1751 TTTCCAGAC TTTATCCTGG AGCCCAAAGA GCTGGATGAT GTTTATGACG
1801 GGTACGAAAT TTCTGAAGAT TCTTTCTTCC AAAACATGTT GAATTTGTAC
1851 AACTTCTCTG CCAAGGTTAT GGCTGACCAG CTCCGCAAGC CTCCAGCCG
1901 AGACCACTGG AGCATGACCC CCCAGACAGT GAATGCCTAC TACCTTCCAA
1951 CTAAGAATGA GATCGTCTTC CCCGCTGGCA TCCTGCAGGC CCCCTTCTAT
2001 GCCCGCAACC ACCCCAAGGC CCTGAACTTC GGTGGCATCG GTGTGGTCAT
2051 GGGCCATGAG TTGACGCATG CCTTTGATGA CCAAGGGCGC GAGTATGACA
2101 AAGAAGGGAA CCTGCGGCCC TGGTGGCAGA ATGAGTCCCT GGCAGCCTTC
2151 CGGAACCACA CGGCCTGCAT GGAGGAACAG TACAATCAAT ACCAGTCAA
2201 TGGGGAGAGG CTCAACGGCC GCCAGACGCT GGGGGAGAAC ATTGCTGACA
2251 ACGGGGGGCT GAAGGCTGCC TACAATGCTT ACAAAGCATG GCTGAGAAAG
2301 CATGGGGAGG AGCAGCAACT GCCAGCCGTG GGGCTACCA ACCACCAGCT
2351 CTTCTTCGTG GGATTTGCCC AGGTGTGGTG CTCGGTCCGC ACACCAGAGA
2401 GCTCTCACGA GGGGCTGGTG ACCGACCCCT ACAGCCCTGC CCGCTTCCGC
2451 GTGCTGGGCA TTCTCTCCAA CTCCCGTGAC TTCCTGCGG ACTTCGGCTG
2501 CCCTGTCGGC TCCCCATGA ACCCAGGGCA GCTGTGTGAG GTGTGGTAGA
2551 CCTGGATCAG GGGAGAAATG CCCAGCTGTC ACCAGACCTG GGGCAGCTCT
2601 CCTGACAAAG CTGTTTGCTC TTGGGTTGGG AGGAAGCAAA TGCAAGCTGG
2651 GCTGGGTCTA GTCCCTCCCC CCCACAGGTG ACATGAGTAC AGACCCTCCT
2701 CAATCACCAC ATTGTCGCTC TGCTTTGGGG GTGCCCCCTG CTCCAGCAGA
2751 GCCCCACCA TTCACTGTGA CATCTTTCCG TGTCACCCTG CCTGGAAGAG
2801 GTCTGGGTGG GGAGGCCAGT TCCATAGGA AGGAGTCTGC CTCTTCTGTC
2851 CCCAGGCTCA CTCAGCCTGG CGGCCATGGG GCCTGCCGTG CCTGCCCCAC
2901 TGTGACCCAC AGGCCTGGGT GGTGTACCTC CTGGACTTCT CCCCAGGCTC
2951 ACTCAGTGCG CACTTAGGGG TGGACTCAGC TCTGTCTGGC TCACCCTCAC
3001 GGGCTACCCC CACCTCACCC TGTGCTCCTT GTGCCACTGC TCCAGTGCT

```

FIGURE 1A

Dock t No.: CL001200-DIV II
Serial No.: TO BE ASSIGNED
Inventors: Ming-Hui WEI et al.
Title: ISOLATED HUMAN ZINC METALLOPROTEASES...

3051 GCTGCTGACC TTCCTGACA GCTCCTAGTG GAAGCCCAAG GGCCTCTGAA
3101 AGCCTCCTGC TGCCCACTGT TTCCCTGGGC TGAGAGGGGA AGTGCATATG
3151 TGTCGCGGT ACTGGTTCCT GTGTCTTAGG GCACAAGCCT TAGCAAATGA
3201 TTGATTCTCC CTGGACAAAG CAGGAAAGCA GATAGAGCAG GGAAAAGGAA
3251 GAACAGAGTT TATTTTACA GAAAAGAGGG TGGGAGGGTG TGGTCTTGGC
3301 CCTTATAGGA CCTGTGCCA ATAAACAGAC ATGCATCCGT CAAAAA
3351 AAAAAA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-113
Start Codon: 114
Stop Codon: 2547
3'UTR: 2550

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 18000005141003 /altid=gi 7662200 /def=ref NP_055508.1 KIAA...	1550	0.0
CRA 18000005084162 /altid=gi 2136744 /def=pir I46078 endotheli...	1430	0.0
CRA 18000005012182 /altid=gi 1706565 /def=sp Q10711 ECE2_BOVIN ...	1430	0.0
CRA 150000075554683 /altid=gi 9789315 /def=gb AAF98287.1 AF2302...	1027	0.0
CRA 1000682324124 /altid=gi 5821116 /def=dbj BAA83687.1 (AB031...	1001	0.0
CRA 108000024636251 /altid=gi 12721007 /def=ref XP_001827.2 en...	1001	0.0
CRA 18000004932659 /altid=gi 627989 /def=pir A53679 endothelin...	1000	0.0
CRA 18000005060029 /altid=gi 1706564 /def=sp P42893 ECE1_RAT EN...	996	0.0
CRA 18000005155376 /altid=gi 3287157 /def=emb CAA19767.1 (AL03...	995	0.0
CRA 18000004985166 /altid=gi 1082351 /def=pir JC2521 endotheli...	995	0.0

BLAST dbEST hits:

	Score	E
gi 6837875 /dataset=dbest /taxon=9606 /org=...	1094	0.0
gi 5925169 /dataset=dbest /taxon=9606 ...	983	0.0
gi 10725997 /dataset=dbest /taxon=96...	519	e-144
gi 2162041 /dataset=dbest /taxon=9606 ...	400	e-109

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|6837875 lung
gi|5925169 amygdala
gi|10725997 adrenal gland
gi|2162041 total fetus

From tissue screening panels:

Hippocampus

FIGURE 1B

```

1 MNVALQELGA GSNMVEYKRA TLRDEDAPET PVEGGASPD MEVGKGASPF
51 SPGPSPGMTP GTPRSSGLFW RVTCPLHRSI SGLCSRTMVG FQKGTRQLLG
101 SRTQLELVLA GASLLLAALL LGCLVALGVQ YHRDPSHSTC LTEACIRVAG
151 KILES LDRGV SPCEDFYQFS CGGWIRRNPL PDGRSRWNTF NSLWDQNQAI
201 LKHLLENTTF NSSSEAEQKT QRFYLSCLQV ERIEELGAQP LRD LIEKIGG
251 WNITGPWDQD NFMEVLKAVA GTYRATPFFT VYISADSKSS NSNVIQVDQS
301 GLFLPSRDY LNR TAN EKVL TAYLDYMEEL GMLLGGRPTS TREQM QQVLE
351 LEIQLANITV PQDQRDEEK IYHKMSISEL QALAPSMDWL EFLSFLLSPL
401 ELSDSEPVVV YGMDYLQQVS ELINRTEPSI LNNYLIWNLV QKTTSSSLDRR
451 FESAQEKLE TLYGTTKSCV PRWQTCISNT DDALGFALGS LFVKATFDRQ
501 SKEIAEGMIS EIRTA FEEL GQLVWMDEKT RQAAKEKADA IYDMIGFPDF
551 ILEPKELDDV YDGYEISEDS FFQNMNLNLYN FSAKVMADQL RKPPSRDQWS
601 MTPQTVNAYY LPTKNEIVFP AGILQAPFYA RNHPKALNFG GIGVVMGHEL
651 THAFDDQGRE YDKEGNLRPW WQNESLA AFR NHTACMEEQY NQYQVNGERL
701 NGRQTLGENI ADNGGLKAAY NAYKAWLRKH GEEQQLPAVG LTNHQLFFVG
751 FAQVWC SVRT PESSHEGLVT DPHSPARFRV LGTLSNSRDF LRHFGCPVGS
801 PMNPGQLCEV W (SEQ ID NO:2)

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION

N-glycosylation site

Number of matches: 9

```

1      207-210 NTTF
2      211-214 NSSS
3      252-255 NITG
4      312-315 NRTA
5      357-360 NITV
6      424-427 NRTE
7      580-583 NFSA
8      673-676 NESL
9      681-684 NHTA

```

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE

cAMP- and cGMP-dependent protein kinase phosphorylation site

18-21 KRAT

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 8

```

1      21-23 TLR
2      62-64 TPR
3      220-222 TQR
4      272-274 TYR
5      340-342 STR
6      465-467 TKK
7      582-584 SAK
8      757-759 SVR

```

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site

Number of matches: 19

```

1      21-24 TLRD
2      30-33 TPVE
3      103-106 TQLE
4      161-164 SPCE

```

FIGURE 2A

5 192-195 SLWD
6 212-215 SSSE
7 214-217 SEAE
8 314-317 TANE
9 340-343 STRE
10 376-379 SISE
11 398-401 SPLE
12 403-406 SDSE
13 445-448 SSLD
14 453-456 SAQE
15 478-481 SNTD
16 514-517 TAFE
17 613-616 TKNE
18 705-708 TLGE
19 763-766 SSHE

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

Number of matches: 4

1 365-372 RRDEEKIY
2 457-463 KLEETLY
3 535-542 KEKADAIY
4 555-561 KELDDVY

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 14

1 9-14 GAGSNM
2 57-62 GMTPGT
3 61-66 GTPRSS
4 122-127 GCLVAL
5 159-164 GVSPCE
6 271-276 GTYRAT
7 331-336 GMLLGG
8 335-340 GGRPTS
9 464-469 GTKKSC
10 643-648 GVVMGH
11 714-719 GGLKAA
12 715-720 GLKAAAY
13 782-787 GTLSNS
14 795-800 GCPVGS

[7] PDOC00047 PS00048 PROTAMINE_P1
Protamine P1 signature

776-787 ARFRVLGTLSNS

[8] PDOC00129 PS00142 ZINC_PROTEASE
Neutral zinc metalloproteinases, zinc-binding region signature

645-654 VMGHELTHAF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	43	63	0.638	Putative
2	109	129	2.142	Certain
3	380	400	0.619	Putative
4	736	756	0.890	Putative

FIGURE 2B

Docket No.: CL001200-DIV II
Serial No.: TO BE ASSIGNED
Inventors: Ming-Hui WEI et al.
Title: ISOLATED HUMAN ZINC METALLOPROTEASES...

BLAST Alignment to Top Hit:

>CRA|18000005141003 /altid=gi|7662200 /def=ref|NP_055508.1| KIAA0604
gene product [Homo sapiens] /org=Homo sapiens
/taxon=9606 /dataset=nraa /length=765
Length = 765

Score = 1550 bits (3969), Expect = 0.0

Identities = 765/811 (94%), Positives = 765/811 (94%), Gaps = 46/811 (5%)

Query: 1 MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEVGKGASPFSPGSPGMTP 60
Sbjct: 1 MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEV----- 43

Query: 61 GTPRSSGLFWRVTCPLHRSISGLCSRTMVGFQKQTRQLLGSRTQLELVLAGASLLLAALL 120
GFQKQTRQLLGSRTQLELVLAGASLLLAALL
Sbjct: 44 -----GFQKQTRQLLGSRTQLELVLAGASLLLAALL 74

Query: 121 LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESIDRGVSPCEDFYQFSCGGWIRRNPL 180
LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESIDRGVSPCEDFYQFSCGGWIRRNPL
Sbjct: 75 LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESIDRGVSPCEDFYQFSCGGWIRRNPL 134

Query: 181 PDGRSRWNTFNSLWDQNQAILKHLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP 240
PDGRSRWNTFNSLWDQNQAILKHLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP
Sbjct: 135 PDGRSRWNTFNSLWDQNQAILKHLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP 194

Query: 241 LRDLEIKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTUYISADSKSSNSNVIQVDQS 300
LRDLEIKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTUYISADSKSSNSNVIQVDQS
Sbjct: 195 LRDLEIKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTUYISADSKSSNSNVIQVDQS 254

Query: 301 GLFLPSRDYYLNRTANEKVLTAFLDYMEELGMLLGGRRPTSTREQMQQVLELEIQLANITV 360
GLFLPSRDYYLNRTANEKVLTAFLDYMEELGMLLGGRRPTSTREQMQQVLELEIQLANITV
Sbjct: 255 GLFLPSRDYYLNRTANEKVLTAFLDYMEELGMLLGGRRPTSTREQMQQVLELEIQLANITV 314

Query: 361 PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS 420
PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS
Sbjct: 315 PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS 374

Query: 421 ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLETLTYGTTKSCVPRWQTCISNT 480
ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLETLTYGTTKSCVPRWQTCISNT
Sbjct: 375 ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLETLTYGTTKSCVPRWQTCISNT 434

Query: 481 DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFFEEALGQLVWMDEKTRQAAKEKADA 540
DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFFEEALGQLVWMDEKTRQAAKEKADA
Sbjct: 435 DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFFEEALGQLVWMDEKTRQAAKEKADA 494

Query: 541 IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMNLNLYNFSAKVMADQLRKPPSRDQWS 600
IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMNLNLYNFSAKVMADQLRKPPSRDQWS
Sbjct: 495 IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMNLNLYNFSAKVMADQLRKPPSRDQWS 554

Query: 601 MTPQTVNAYYLPKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE 660
MTPQTVNAYYLPKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE
Sbjct: 555 MTPQTVNAYYLPKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE 614

Query: 661 YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY 720
YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY
Sbjct: 615 YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY 674

FIGURE 2C

Docket No.: CL001200-DIV II
 Serial No.: TO BE ASSIGNED
 Inventors: Ming-Hui WEI et al.
 Title: ISOLATED HUMAN ZINC METALLOPROTEASES...

Query: 721 NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV 780
 NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV
 Sbjct: 675 NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV 734

Query: 781 LGTLSNSRDFLRHFGCPVGSPMNPGLCEVW 811
 LGTLSNSRDFLRHFGCPVGSPMNPGLCEVW
 Sbjct: 735 LGTLSNSRDFLRHFGCPVGSPMNPGLCEVW 765 (SEQ ID NO:4)

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
CE00310	E00310 neutral_endopeptidase	456.3	2.6e-133	1
PF01431	Peptidase family M13	270.4	2.4e-77	1
CE00339	E00339 vitamin_D_receptor	4.3	1.1	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00339	1/1	367	398 ..	412	443 .]	4.3	1.1
PF01431	1/1	607	810 ..	1	225 []	270.4	2.4e-77
CE00310	1/1	140	811 .]	66	798 .]	456.3	2.6e-133

FIGURE 2D

Docket No.: CL001200-DIV II
Serial No.: TO BE ASSIGNED
Inventors: Ming-Hui WEI et al.
Title: ISOLATED HUMAN ZINC METALLOPROTEASES...

```
1  NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
51 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
1001 NNNNNNNNNN NNNNNNNNNN NNNNNNNACC TTAGACTTGA CAGGCCTGCT
1051 TAGTCGGA CTAAAGCACC CCTTTGCTTT TCGTTAAATA TTGCTTGGTG
1101 TTAGTTTTTT TTCTCCTTGT AAATCTCCCA AATAAACGG TTTGCTTCC
1151 CCAAGTTAGA AGTGTTAGCA CGTCTTTTCT TTAATATCT GTGCATGGCT
1201 GTTTTTTTCC CTGCCAATTT GTCACCATCT GTAACCCTCC CTTTATGAGA
1251 CGATCTGATG ACAGCAGTTA TCTTGAGAG TAGAAGTGTG GTCTGAAGC
1301 GCCATGGAAG AGTAGAGTCA GTGTATGCTG TGTGTGTGTG GAGTGTATGC
1351 TCCCCCTGCA CTTGGTGTGT GTACATACAG AAACACAGTG TGCGTGTGTG
1401 TTGGCTCTGG GTGTGTGTGT CGTGTGTACA CTGTGTGTGA GTATGCAGTG
1451 TGTGTACATT CTGTGGGCAT CTCGTGTGTG TGTGGAAGTG GTGCTGGGCG
1501 TCGTGCCCTG CCGTGTCCTT GCGCGCTTGG CGTCTATGCG TTCTCTGCAC
1551 ATAGGTAGGT ACCACGTGCA CACCCTGAAT GTGAGTGAAC TGCTGTGTG
1601 CTATGTATTT GCCGGCTGAA GAGGGGCTGT GTGGACTACT GGGGAAGAC
1651 GTTCTCANG AGGGCATAAT TTCTCTAAAG TGCTTAAAGG GGATGGAGAG
1701 AGCCTGAAAT TTGGGGGAAG TAGGCCAAGG AGTATTATCA ACGTCTGGGC
1751 CTGGTTGAAT TTCATTACTT TTCTTAGGAA AGTAAATTAT GGGTGGCTTG
1801 AAGGAGGGTG CTGCTGAGAT GGGGGGCGGA CCATGAAGCG TGAGGGGGTC
1851 TCCGGTGTG CTGGAGGGCA GCTGGAGCCT GCGGAGAGCC TCGGCGCGCT
1901 CCTCCTCTC CCCACCCCTC CCCACCCCTC GGGCGGGGCT CCGCGTGGGG
1951 CCGTGGAATC GGGCGGGGGG GGGGCGGGCC GCGGCGGAGC GGGGTGTGTG
2001 CGCGGCGGCC GTGATGGCTG GTGACGGCGG GGCCGGGCGG GGGACGGGGG
2051 CCGCGGCGCG GGAGCGGGCC AGCTGCGGG AGCCCTGAAT CACCGCCTGG
2101 CCCGACTCCA CCATGAACGT CGCGCTGCAG GAGCTGGGAG CTGGCAGCAA
2151 CGTGAGTGGG GGCCCCGGGC TCCACGGGAG GGGACTGGGT GGAGGGGGAC
2201 GAGGAGAGG GGTGCGCGCG GGAGGGGCG GCGGTGCCCC GCTCGCGGAG
2251 GTAAGGCTGC CTCCGGGGCC TGGTGGAGGG GTGATAGAGA GACCCGGGGC
2301 CCGAGAGCAG GGCAGGTGGG AAGGGAAGGG CCCTCTTAGC AGGGCGGAGG
2351 GGTCCGCGAG GCAGGGAGCA CTGGGGCAGG GTCGTGGGCA AATAGCCCTC
2401 TCTGCCCTGAC CTGCGTTGGC AACCCCGACT GTCTGGCAGA TGGTGGAGTA
2451 CAAACGGGCC ACGTTCGGG ATGAAGACGC ACCCGAGACC CCCGTAGAGG
2501 GCGGGGCTC CCCGACGCC ATGGAGGTG GCAAGGGGGC TTCCCTTTC
2551 TCACCAAGCC CCAGCCTGG CATGACGCTT GGCACACCCA GGAGCTCTGG
2601 GCTGTTCTGG AGGGTCATCT GCGCCACCT CCGCTCCATC TCTGGCCTCT
2651 GCTCTAGGAC TATGGTGAGG CGATGCTAAG CCGTGACGTT GCACAAAACA
2701 GACTCAAGGC TCAACTCACT GGCTGGCCTC ATTGCCCCCG GGCCAGAGT
2751 TAACCTGTG GCTCTGAAAA CTGCTGTGG CTTACCCCTC TGGTAATCTT
2801 GGATCCCTGC CTGTCATCTC AGTCACTCTC TGTCCCCCTG TGTCCCCAG
2851 GTGGGATTCC AGAAGGGGAC AAGACAGCTG TTAGGCTCAC GCACGAGCT
2901 GGAGCTGGTC TTAGCAGGTG CCTCTCTACT GCTGGCTGCA CTGCTTCTGG
2951 GCTGCCCTGT GGCCCTAGGG GTCCAGTACC ACAGAGGTAG GTGGGCCAC
3001 ACTCTTCGTC AGTATTCATA ACTAGGGGTT CTGAGGCCT AAGGCCTCT
```

FIGURE 3A

```

3051 AAGATTTTCA CTTGTGGGAA CCAAGCCTTC CCTGCAGAAA AGCCCCCGGC
3101 TTTGCTTTCT CTTCCCAACC TTCCTGCTGT CATGGCCCTT GCAGAGTTTG
3151 CCTCTTCCAG ACAGACAGAC TGACAGTCTC CTACCCCTCG GCCATGTTCC
3201 CTACCACAGA CCCATCCAC AGCACCTGCC TTACAGAGGC CTGCATTCCA
3251 GTGGCTGGAA AAATCCTGGA GTCCCTGGAC CGAGGGGTGA GCCCCGTGA
3301 GGACTTTTAC CAGTTCTCCT GTGNNNNNNN NNNNNNNNNN NNNNNNNNNN
3351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3451 NNNNNNNNNC TTAGCAAATA GGCAGTGTCC CATGAATGAG GAAGTGGATG
3501 GTTCTGTGAA CACTCCCAGA GGGTGGGGAG GCAGAGAGCA GGGGACTATT
3551 GAGAAGTGCA GATGGGTTTG ATGGGGGCAG AACTCTGGGT ACAATGGAGG
3601 GCCGCTTCTC TGCACCTCTG TTGGAGCACT GTCGTGGTGT GGTAGACACC
3651 AGGGAGCCTG TACTGCTTAG ATATCCTTGG GTCTCCATGG ACAGGGAGAG
3701 GAAGCCACGG CTTGCTGTTT CAGACACTCT TCCTGGGTCT GCGTTAGCAG
3751 GACTGCTCAT TGACAAGGCA AGGAGAGAAA CCGAGCAAGG GCCAGGGACT
3801 CCCCCTCAGC AGTTAACGTA ATTGCCACCT GGATCCTGTG TTCTGCCCCA
3851 CAGAAAACAC CACCTTCAAC TCCAGCAGTG AAGCTGAGCA GAAGACACAG
3901 CGCTTCTACC TATCTTGCCT ACAGGTGGAG CGCATTGAGG AGCTGGGAGC
3951 CCAGCCACTC AGAGACCTCA TTGAGAAGGT AGGGCCACTG AGCCGGTTGA
4001 GGGCAGGGGA GCAGGAGAGG CCTTGAGAGA GGAGATGGCC CAGGAACGCT
4051 TTGGGAGCTC CTGCACTAAT CATTCCACTT ATGGTCTCTA CATAGATTGG
4101 TGGTTGGAAC ATTACGGGGC CCTGGGACCA GGACAACCTT ATGGAGGTGT
4151 TGAAGGCAGT AGCAGGGACC TACAGGGCCA CCCCATCTT CACCGTCTAC
4201 ATCAGTGCCG ACTCTAAGAG TTCCAACAGC AATGTTATCC AGGTGATGAG
4251 CTGGGAAAGG GTGGGAGAG ACTTAGGGAC ACTTTGCTGA GCCCAGACTT
4301 CCCTCTCCTG TGACAGGCAG GCTGGGCTGA CCCCCCGGCC CCACCCCTAC
4351 CCCCCTCGG GAATTCAGGT TCCCATGGTG GGGAAAGCGA GGGGCTCACC
4401 TCCTTTCTCT GACATTGCAG GTGGACCAGT CTGGGCTCTT TCTGCCCTCT
4451 CGGGATTACT ACTTAAACAG AACTGCCAAT GAGAAAGTAA GGAACATCTT
4501 CCGAACCCCT ATCCCTACCC CTGGCTGAGC TGGGCTGATC CCTGTTGACT
4551 TTTCCTTTTG CCAAGGGTCA GAGCAGGGAA GGTGAGCCTA TCCTGTCACC
4601 TAGTGAACAA ACTGCCCTCT CTTTCTTTCT TCTTTTCTTC CTCCCTCCCT
4651 CCCTTTCTTC CCCTTTCTCT TCCTTCCTTC CTCTTATTCT TCTAGTAGGT
4701 TTCATAGACA CCTACTGTGT GCCAGGTCCA GTGGGGGAAT TCTGAGATAT
4751 AAGTTTNCAG AGCCCATTGC CAGCAGGAGA GGGGATCCTT TAGAGTCGCA
4801 CAAACAGGTC AGTCAAGTCT AAAGACNNNN NNNNNNNNNN NNNNNNNNNN
4851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5101 NNNNNNNNNN NNNNNNNNNN GCCTGNACTT GCATGCACCG CGGTTCGGCT
5151 NCTAGNAGNA TCCCCCACTT GCACTCCAGC CTGGGTGACG GAGAGAGACT
5201 CCGACTCAAA AAAAAAAGAG AAAAAAGAAAG AAAAAAGAAAG AAGGAACAGT
5251 TTAAACAAAA GTGTTGATGA GGCTGAGCAC AGTGGCTCAC ACCTGTAATC
5301 CCCGCACTTT GGGAGGCTGA GGCCGGCGGA TCACTTGAGG TTAGGAGTTC
5351 AAGACCAGGC TGGCCTACAA GGTGAAAACC CGTCTCTACT AAAAATACAA
5401 AAATTAGCCA GGCATGGTGG TGTGCACCTG TAATCTCAGC TACTTGGGAG
5451 GCTGAGGCAA AGAGAATCGC TTGAATCCAG GAGGCAGAGG TTGCAGTGAG
5501 CTGAGATGGC ACCACTGCAC TCCAGCCTGG GCAACAGAAC AAGACTTCAT
5551 CTCAAAAAAA AAAAAAAGAG TGTGACGAG GGAAAGGCTA GGTGTGTCTG
5601 GACCATGGCA AGGGGTCCAC TGTGGTAAAA TATAGAACTC AAGGCAGATG
5651 AGAGGCTGGA GAGGTGGGCA GGAATGGGTT ATGGAGGGGA CCTTGAATAG
5701 CACACTACGG AGTTTATTCT GTAGCTCCCG GAGAGCCATT GCATGCTCCA
5751 AAGTAGGGAG GGAGCGCANT GCTTTGGGAA GTCAGTTTGT TTGGGGTGTG
5801 AAGAGTANAT GTGAGAACNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

```

FIGURE 3B

9151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
9951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
10951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
11401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNAG	TTCCAGGCCC
11451	ACCCTTGGGC	CAAACATGTT	GAAGACCGCC	ATGCTGTAGC	TAGAACTTAC
11501	AAAAGATGTA	AGCCTGGGCA	TAGGTGGCCG	GGTGCCGTG	TGGTCGCCAC
11551	GCTATCTTGG	GGAGGGATTA	AGGGCAAGGA	AAATTCACCT	TGAGGCCCAA
11601	GGAAGGCACA	AGGGTTATCA	CGTGAAGCCG	AGGATCACCA	TCACCATGCA
11651	CTAACACGCC	TTGGGCAAGC	ACGAAGCGAG	GAGTTGCCAT	CTCAAACAA
11701	AAACGAAAAA	CAAACAAACA	AAATGCTAAT	CAACTGTCAT	TGGTAAGGCT
11751	TCTGGTCAAC	AGTATGCTGT	CAATAGTTAA	GTTTTTGGGC	TGGGCGCAGT
11801	GGCTCACGCC	TGTAATCCCA	GCACTTTGGG	AGGCCAAAGC	GGGTAGATCA
11851	CCTGAGGTCA	GGAGTCGAGA	CTAGCCTGGC	CAACATGGCG	AAACCCAGTC
11901	TCTACTAAAA	ATATAAAAT	TAGCCAGGCG	TGGTGGTGGG	CACTTGTAAT
11951	CCCAGTACT	CAGGAAGCTG	AGGCAGAACT	GCTTGAAGTG	GGAAGTGGAG
12001	GTTGCAGTGA	GCCGAGATCG	TGCCATTGCA	TTCCAGCCTG	GGCGACAAGA
12051	GCAAACTCC	ATCTCAAAAA	AAAAAAGTT	GTTTTTGGGG	
12101	AGTCAAAAT	GAGGCCAGGC	GCAGTGGCTC	ATGCCTGTAA	TCACAGCACT
12151	TTGGGAGGCC	GAGGCGGGTG	GATCACCTGA	GGTCAGGAGT	TCGTGACCAG

FIGURE 3D

```

12201 CTTGGCCAAC CTGGTGAAAC CCCGTCTCTA CTAAAAATAC AAAAATTAGC
12251 CGGGCATGGT GCGGGGCGCC CGTAATCTCA GCTACTTGGG CGGCTGAGGC
12301 AGGAGAATTG CTTCAACCCG GGAGGCAGAG GTTGCAATGA GCTGAGATCG
12351 CGCCACTGCA CTCCAGCCTT GCGGACAGAG GGAGACTCCA TGTCAAATTA
12401 AAAAAAGAC CCCAGGATTT TGGACTGTGC AGGGGTGCGT GCGGCAAAACC
12451 CCCACGTTGT TCAAGGTCAA CTGTACACTG TCATAGTCGG GAAAACTTCA
12501 TCACTGCAGC TGCTCCTGTT TCTTGAAACC TGAAGCGGGA AACTGGATCC
12551 TGGGACACTA CTGCCCCCTA TCGCCTGTTG GTCTTCAAAG AAATAATCCC
12601 TTCAATTTTG CAAGGCCTGT GGTGTCATTC CCTTTTAAAC GATAAGGAAA
12651 CCGAGGCCAG GACGTGGTGG AAAATAATCA AGGTCACACA TCTATGTGCA
12701 AAAGTGGAGT AACAACCCAG GCTCCTCATT CCCAGGTGAC TCCAGTGACC
12751 TCAATTGACA TGAAATGTGT GAGGTCCTTC TGTGGCCCTG TGGCAGGGCC
12801 TGAAGAGGAC AGCGTATGTA AATCAAGTCT TGTGCCTTCA TGAGTGAGGC
12851 AGAGTAGAAA ATAACAGTAA TTTACTAGGA CCGAATCTGC ATTGTAAACA
12901 GAGAGGAAAG GGCTAGTATT TGGCAGAAAG ATGTCAAGGA ACATTTTAGA
12951 GATAAGAGGT GACATTTGGG TTCTGAGGGA TGAGTAGGAG TGTGCCAGGG
13001 TGCAAAGGAT GAAAAGACAG CTCTAGCAGC TGGTAAGGGC TAAGGGGCAT
13051 GGAGAAACAG CAAGACTTTG GGGAACTGGT AGAATTTCTA TTCTGGAAAA
13101 TTTGAACAAG GTAATTTTGT GTGTGTGGTT AAGGTATTAC ATACATACAG
13151 TAAAAATAAA TGCAATAGTT GCTGGGTGTG GAGGCTCAGC CCTGTTAATC
13201 CCAGTACTTT GGAAGGCAGA GCGGGGTGGA TCATCTGAAG GTCAGGAGTT
13251 CGAGACCAGC CTGACCAACA TGGTGAAAAC CCGTCTCTAC TAAAAATACA
13301 AAAATTACCT GGGTGTGGTG GCAGGCGCCC GTAATCCCAG CTACTTGGGA
13351 GGCTAAGGGA GAAGAATAGC TTGAAACCCG GAGGTGGAGG TTGCAGTGAG
13401 CTGAGATTGC ACTATTGCGC TCCAGCCTGG GTGACAAGAG TGAAAAGCTG
13451 TCTCAAAATA AAATAAAAT GTAAATAGTCT AATTGATTTT TTTAAAAAAT
13501 GTAGACATCC ACGTATCTAC CACCTAGGTA AAGATACTAG AGATTCCAGC
13551 AACCTGGGAG GATCCCTCGT GCGGCTTTCA GGTCTATATG AGCCTCCACC
13601 GTTCCCCAGT CCCCTGGAAG GAGAGGGGGT GGGAGAGGCA ACATGAAACC
13651 TAAAAACCAG TGGGCTTCGC GCCTGTAATC CCAGCTATTG GGTTGGCTGA
13701 GGCAGGAGGA TCACTTGCCC AGGAGTTGGA GGCTGCAGTG AGCTATGATC
13751 GCGCCACCCG ACTCCAGCCT GGGCGACAGA TCAAGACCCC ATCTCTAAGC
13801 AAACAAACAA ATAAACACCC CTCAAACCCC ATGGCTTCAG GCCTGGCGCG
13851 GTAGCTTACT TCTGTAATCT CAGCACTTTG GGAGGCCGAG GAGGGCGGAT
13901 CACTTGAGGT CAGGAGTTCC AGACCAGACT GGCCAACATG GCGAAACCCC
13951 GTCTCTACTA AAAAATAAAA AAAAAAAAAA ATTGGCCGGG CGCGGTGGCT
14001 CACACCTGTA ATTACCAGCA GNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14051 NNNNNNNNNN NNNNNNNNNN NTTTTAAAGA ATGTAGACAT CCACGTATCT
14101 ACCACCTAGG TAGAGATACT AGAGATTCCA GCAACCTGGG AGGATCCCTC
14151 GTGCGCCTTT CAGGTCTATA TGAGCCTCCA CCGTTCCCCA GTCCCCTGGA
14201 AGGAGAGGGG GTGGGAGAGG CAACATGAAA CCTAAAAACC AGTGGGCTTC
14251 GCGCCTGTAA TCCAGCTAT TGGGTTGGCT GAGGCAGGAG GATCACTTGC
14301 CCAGGAGTTG GAGGCTGCAG TGAGCTATGA TCGCGCCACC GCACTCCAGC
14351 CTGGGCGACA GATCAAGACC CCATCTCTAA GCAAACAAAC AAATAAACAC
14401 CCTCAAAAC CCATGGCTTC AGGCCTGGCG CGGTAGCTTA CTTCTGTAAT
14451 CTCAGCACTT TGGGAGGTCA AGGTGGGCGG ATCACTTGAA GTAAGGAGTT
14501 CAAGTACCAT CCTGGCTAAC ACGGTGAAAC CCCGTCTCTA CTGAAAAGAC
14551 AAAAAATTTA CCGGGGCGTG GTGGCGGGCG CCTTTAGTCT CAGCTACTCG
14601 GGAGGCTGAG GCAGGAGAAT GGCGTGAACC CGGGAGGTGG AGCTTGCACT
14651 GAGCTGAGAT CGCACCACGT CACTCCAGTG TGGGTGACAG AGTGAGACTC
14701 CATCTCAAAA AAAAAAAAAA AGAAGTCAAA GTAGTAGAAA CTGCTGATAG
14751 ACTGAATGTG GGGGGTTAGG GAGATGGAGG AAGCTGAGTG ACTCCCAGGT
14801 TTCTTGCATG GGGGACTGAC TGGATATAAA ATTAGTTGTG GGCCGGGCAC
14851 GGTGGCTCAT GCCTTTAATC CCAGCACTTT GGGAGGCCAA AGCGGGCAGA
14901 TCACTTGAGC TCAGGAGTTC AAGACCAGCC TGGGAAACAT GGTGAGACCC
14951 CTTCTGTAAG GNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
15001 NNNNNNNNNN NNNNNNNNNN TTTTGGCTC TGNTCGGTCA CTAGCANGCA
15051 AGTTATTGGG AGTCTACAAG ATTCTTTCAC ACTATGCCCT CAAAATTGAC
15101 TGTTTCATGTA TGTGCAGACA TATAGAAAAA CAACGGGAGC CAGGCGCGGT
15151 GGCTCACGCC GGTAATCCCA GCACTTTGGG AGGCCAAGGC GGGTGAATCA
15201 TGGGGTCAGG AGTTCGAGAC CAGCCTGGCC AACATGGTGA AACCTGGTCT

```

FIGURE 3E

15251	CTACTAAAAA	TACAAAAAAT	TAGCCGGGCG	TGGTGGCGGG	TGTCTGTAAT
15301	CCCAGCTACT	TGGGAGGCTG	AGGCAGGAGA	ATCACTTGAA	CCCAGGAGGC
15351	GGAGGTTGCA	GTGAGCCGAG	ATCGCGCCAG	TGCACTCCAG	CCTGGGCGAC
15401	AGAGCAAAAC	TCTGTCTCAA	AAAAAAAAAA	AAAAAAAAAG	AAAGAAAAGA
15451	AAAACAACCTG	GATGTAAATT	GATGAACAAA	TGAAGTAGTG	CTGCTTTGGG
15501	CAGTGGGATT	ATAAGAGTCC	TTTAAAGTTG	TCTATGTGTT	TATGTTTAAAC
15551	TATATAACTA	GAAGAAATAT	TTATTTATTA	GGATATGATA	ATGGATGTGC
15601	TTAAAGTATT	ACCTGTAAGG	ATGTTTATGG	TTTTTATGGC	AATGTTGTTT
15651	ATAATAGCAG	AAAATGAGAA	CAGGTTAAAT	GTCCAACAT	AGGGTAAAGG
15701	AAAAATAAAT	TGTGGTTAGG	ATGGGTTGTG	AGGATCCTTA	AATGGCTGAT
15751	ATATCTTTCA	GCAAAAAAAG	TAGGTTACAA	AAAATATATA	CCCTATACAA
15801	CATAATCCCA	TATTTTATAT	GCATATCAGG	GGAGGGAAAA	ACTCTAGAAG
15851	TGGGTAATCA	AAATGTTAAA	AGAACTTATC	TATGAATGAG	TGCTTTATAA
15901	CTGGTCTGTT	CTTCAATTCT	CAATTTTCCA	AATTTTCTGT	GAATGTCCTC
15951	TTTTCATAAT	CAGATAAAAA	TCATTGCACT	AGGCTGGGCG	TGGTGGTTCA
16001	CGCTTGTAAT	CCCAGCACCT	TGGGAGGCTG	AGGCGGGTGG	ATCACGTGGT
16051	CAGGAGTTCA	AGACCAACCT	GGCCAAGATG	GTGAAACCCC	AGCTCTACTA
16101	AAAATACAAA	AATTACCCGG	GCATGATGGC	GGGAGCCTGT	AATCCTAGCT
16151	ACTTGGGAGG	CTGAGGCAGG	AGAATCGCTT	GAACTCGGGA	GGCGGAGGTT
16201	GCAGTGAGCC	GAGATTGCGC	CACTGCACCT	CATCCTAGGT	AACACAGCCA
16251	GACTCTGTCT	CAAAAAAATA	AAAAAATCAT	TGCACTATAT	TAAATTATAA
16301	TATAATTTGA	TGAACCTTAT	GTCAATTAAA	ATGTGTACTT	AATTAAGAAA
16351	AAAGCCAGCC	ACAATCCCAG	TACCTTTACA	AATGGTGTTC	CCTTCTCATC
16401	GTCTCCAGGT	GCTCAGCCGT	ATTTCTTTAG	TCTAGACGTT	CCCATTTCCT
16451	CTGGGTGGAC	AGGGATGGGG	CACCAAGGGT	GGATGGGTGG	GGCAGGGATG
16501	CATTCACTGC	AGGGGAAGGC	TGACTTTACC	TCCTCCCTCC	CAGGCAGAGG
16551	GGATGATCAG	CGAAATCCGG	ACCGCATTTG	AGGAGGCCCT	GGGACAGCTG
16601	GTTTGGATGG	ATGAGAAGAC	CCGCCAGGCA	GCCAAGGAGA	AAGTGAGCGG
16651	TGGCTAGGGT	TGGGGCGCCA	TCTTGAGGTG	GGGTTCAAGG	ATACAGTTTT
16701	GCTAGGAACC	TGGGGAAGGA	AACAAACCCT	TAACCTGGTC	TCTTCAGGCA
16751	GATGCCATCT	ATGATATGAT	TGGTTTCCCA	GACTTTATCC	TGGAGCCCAA
16801	AGAGCTGGAT	GATGTTTATG	ACGGGGTGAG	TACCTACGCT	CATCAGTACT
16851	GAACCTCAGC	CCTGTAGAGG	GCACTGTTCC	CTGGGCTTAG	AAATTGGGGC
16901	TCAAGCACTG	GGAAAGAGGT	GCTTGTCCGT	TTCTTTTAGA	GGCAGATGGA
16951	GGTAACCAGC	ATTGTTAAAA	TGTTGGCTCT	GTGACAGGCT	GCAGGCCAAA
17001	CAGCAGTGAA	ATATAGTGCT	AACGAGCCAA	GATTTGGAGT	CAAGCCTAAT
17051	CAAATTCTGT	TTCTACCTCT	AACTTTGTAA	CCTTAACAAA	ATCTCTCTAG
17101	GCCTTGGTTT	CATTTTCTGT	AAAATGGGGG	TCCTACTAGT	GCCTTCCTCA
17151	TAGGGTTGTT	GTGAGATAAA	TGAATACAGT	ATGTAAAAAA	ACAGCACCCA
17201	TAACATAAAT	GGCCTTTAAA	TATTGCCAAT	TATGGTTTAC	TAGATATTTT
17251	ACAGTTGAGG	AAACTGAGGT	TTGGAGAGAT	ACTAATGAGT	AGCCAAACTG
17301	GCGCTATTAT	CTTCTCCAAT	GGATTCTCTT	GCTCTCTGTC	TACTTCCCAA
17351	CTTACCACAG	AACAAANNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN

FIGURE 3F

Docket No.: CL001200-DIV II
Serial No.: TO BE ASSIGNED
Inventors: Ming-Hui WEI et al.
Title: ISOLATED HUMAN ZINC METALLOPROTEASES...

```
18301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18801 NNNNNNNNNN NNNNNNNNNN NNNNNNAGAAT CACCAACAGC ATTGGATGAA
18851 AATAAAGAAG AACAAAGAGT TCGTTTGAGA GGAAGCCGGG AAAATTCTCT
18901 CGATAAAGAA ATGCAAGTGC GCGCGCGGCG CAACCACTAC AATAGTGTGT
18951 CGTCCACCCC AGAGAGTGAA GGGGGCCCCC CCCGCCCAA AGGAAAGGGG
19001 TAGTGTCAC GCGCTCCAC AAAGAGAGAG AAGGAAAGAA GTAGTTTTC
19051 CCCCCCGGG GAGAAACCTT GGATGGGGCT CANCCCCCCC TCTTTTTTTT
19101 TCCCGCGAAA ACCCCCCCA AAAAGTTTTT TTTAAAAAAC AAAAAAGGGG
19151 GGTTTGGTTT TTTGGGCCCC GTGGCCCCTT TGGTTTAAAT TGGGAGAAAG
19201 AGGGCTTAAA GGGGGGATT CAGAAAAAAC CCCCCCAA TTGCCCAA
19251 TTGTAATTTC CTAACCCCAA AAGGGGCCCC TAAATTTCC GGGGAAACCC
19301 GTGTGGGCAA TGGCCCATTA GTTTACCCAA TGCCTTTAT GACAAAGGTA
19351 GGGCCCCATG GAGTCGTCCC CTCTAGCCTA GAATCCCAG TGGCTCCTGC
19401 AAGGGCCTTG GGACATTGAT GTAGCCCCAA GGGCCCTGAA GTCTGTGGAC
19451 CAGGGCTGGT GGGGCACTGC TGCCCCAAG AGACGAGCTC TGGTTTTGGT
19501 GGGGTGCAAA GGTGAGTTCT CCTCAGGCG CGAGTATGAC AAAGAAGGGA
19551 ACTGCGGCC TGGTGGCAGA ATGAGTCCCT GGCAGCCTC CGGAACCACA
19601 CGGCCTGCAT GGAGGAACAG TACAATCAAT ACCAGGTCAA TGGGGAGAGG (SEQ ID NO:3)
```

FEATURES:

```
Start: 2113
Exon: 2113-2151
Intron: 2152-2439
Exon: 2440-2664
Intron: 2665-2850
Exon: 2851-2986
Intron: 2987-3209
Exon: 3210-3320
Intron: 3321-3519
Exon: 3520-3636
Intron: 3637-3853
Exon: 3854-3978
Intron: 3979-4095
Exon: 4096-4242
Intron: 4243-4420
Exon: 4421-4486
Intron: 4487-4576
Exon: 4577-4581
Intron: 4582-4707
Exon: 4708-4811
Intron: 4812-5525
Exon: 5526-5591
Intron: 5592-11594
Exon: 11595-11742
Intron: 11743-13150
Exon: 13151-13240
Intron: 13241-16408
Exon: 16409-16642
Intron: 16643-16747
Exon: 16748-16825
Intron: 16826-18962
```

FIGURE 3G

Exon: 18963-19222
Intron: 19223-19321
Exon: 19322-19347
Intron: 19348-19526
Exon: 19527-19647

CHROMOSOME MAP POSITION:
Chromosome # 3

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
2707	A	G	Intron			
4209	C	T	Exon	289	S	S
4355	-	T G	Intron			
15455	-	A C	Intron			

Context:

DNA

Position

2707 TGACCTCGGTTGGCAACCCCGACTGTCTGGCAGATGGTGGAGTACAAACGGGCCACGCTT
CGGGATGAAGACGCACCCGAGACCCCGTAGAGGGCGGGGCCTCCCGGACGCCATGGAG
GTGGGCAAGGGGGCTTCCCCTTTCTCACCAGGCCCGAGCCCTGGCATGACGCCTGGCACA
CCCAGGAGCTCTGGGCTGTTCTGGAGGGTCATCTGCCCCACCTCCGCTCCATCTCTGGC
CTCTGCTCTAGGACTATGGTGAGGCGATGCTAAGCCGTGACGTTGCACAAAACAGACTCA
[A, G]
GGCTCAACTCACTGGCTGGCCTCATTGCCCCGGGGCCAGAGTTAACCTGTGGCTCTGA
AAACTGCCTGTGGCTTACCCTCTGGTAATCTTGGATCCCTGCCCTGCATCTCAGTCACT
CTCTGTCCCCCTGTGTCCCCAGGTGGGATTCCAGAAGGGGACAAGACAGCTGTTAGGCT
CACGCACGCAGCTGGAGCTGGTCTTAGCAGGTGCCTCTCTACTGCTGGCTGCACTGCTTC
TGGGCTGCCTTGTGGCCCTAGGGGTCCAGTACCACAGAGGTAGGTGGGCCACACTCTTC

4209 CCTATCTTGCCTACAGGTGGAGCGCATTGAGGAGCTGGGAGCCAGCCACTGAGAGACCT
CATTGAGAAGGTAGGGCCACTGAGCCGTTGAGGGCAGGGGAGCAGGAGAGGCCCTTGAGA
GAGGAGATGGCCAGGAACGCTTTGGGAGCTCCTGCACTAATCATTCCACTTATGGTCTC
TACATAGATTGGTGGTTGGAACATTACGGGGCCCTGGGACCAGGACAACCTTTATGGAGGT
GTTGAAGGCAGTAGCAGGGACCTACAGGGCCACCCATTCTTCACCGTCTACATCAGTGC
[C, T]
GACTCTAAGAGTTCCAACAGCAATGTTATCCAGGTGATGAGCTGGGAAAGGTGGGGAGA
GACTTAGGGACACTTTGCTGAGCCAGACTTCCCTCTCCTGTGACAGGCAGGCTGGGCTG
ACCCCCCGCCCCACCCCTACCCCGCTCGGGAATTCAGGTTCCCATGGTGGGGAAAGCG
AGGGGCTCACCTCCTTTCTTGACATTGACAGGTGGACAGTCTGGGCTCTTTCTGCCCTC
TCGGGATTACTACTTAAACAGAACTGCCAATGAGAAAGTAAGGAACATCTTCCGAACCC

4355 GAGCTCCTGCACTAATCATTCCACTTATGGTCTCTACATAGATTGGTGGTTGGAACATTA
CGGGGCCCTGGGACCAGGACAACCTTTATGGAGGTGTTGAAGGCAGTAGCAGGGACCTACA
GGGCCACCCATTCTTACCGTCTACATCAGTGCCGACTCTAAGAGTTCCAACAGCAATG
TTATCCAGGTGATGAGCTGGGAAAGGTGGGGAGAGACTTAGGGACACTTTGCTGAGCCC
AGACTTCCCTCTCCTGTGACAGGCAGGTGGGCTGACCCCCCGCCCCACCCCTACCCCT
[-, T, G]
CTCGGGAATTCAGGTTCCCATGGTGGGGAAAGCGAGGGGCTCACCTCCTTTCTTGACAT
TGCAGGTGGACAGTCTGGGCTCTTTCTGCCCTCTCGGGATTACTACTTAAACAGAACTG
CCAATGAGAAAGTAAGGAACATCTTCCGAACCCCATCCCTACCCCTGGCTGAGCTGGGC
TGATCCCTGTTGACTTTCCCTTTGCCAAGGGTCAGAGCAGGGAAGGTGAGCCTATCCTG
TCACCTAGTGAACAACTGCCCTCCTTTCTTTCTTCTTCTCCTCCCTCCCTCCCTT

FIGURE 3H

15455 CACGCCGTAATCCCAGCACTTTGGGAGGCCAAGCGGGTGAATCATGGGGTCAGGAGTT
CGAGACCAGCCTGGCCAAACATGGTGAAACCTGGTCTCTACTAAAAATACAAAAATTAGC
CGGGCGTGGTGGCGGGTGTCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCA
CTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCAGTGCACTCCAGCCTG
GGCGACAGAGCAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAGAAAAGAAAAGAAAA
[- , A, C]
AACTGGATGTAAATTGATGAACAAATGAAGTAGTGCTGCTTTGGGCAGTGGGATTATAAG
AGTCCTTTAAAGTTGTCTATGTGTTTATGTTTAACTATATAACTAGAAGAAATATTTATT
TATTAGGATATGATAATGGATGTGCTTAAAGTATTACCTGTAAGGATGTTTATGGTTTTT
ATGGCAATGTTGTTTATAATAGCAGAAAATGAGAACAGGTAAATGTCCAATATAGGGT
AAAGGAAAAATAAATTGTGGTTAGGATGGGTTGTGAGGATCCTTAAATGGCTGATATATC

FIGURE 3I